Ultrasound-to-Ultrasound Volterra Filter Identification of the Parametric Array Loudspeaker

Chuang Shi and Yoshinobu Kajikawa
Department of Electrical and Electronic Engineering, Kansai University
Osaka, Japan
Email: {r148005,kaji}@kansai-u.ac.jp

***Abstract*—The Volterra filter is a favorable approach to represent a variety of nonlinear systems, including the parametric array loudspeaker (PAL), which is a weak nonlinear acoustic system to create directional sounds. Using the Volterra filter to model the sound process of the PAL saves the computational cost of solving the nonlinear acoustic equation. In the past studies, the Volterra filter is identified from the audio input to the audio output of the PAL, which is known as the audio-to-audio Volterra filter (A2VF). In this paper, the ultrasound-to-ultrasound Volterra filter (U2VF) is recommended. The experiment results validate that the U2VF outperforms the A2VF, in terms of the robustness to the change of the modulation index in the PAL.**

***Index Terms*—parametric array loudspeaker; modulation index; total harmonic distortion; Volterra filter; nonlinear system identification**

I. Introduction

When two finite amplitude waves at close frequencies are transmitted in an extremely narrow beam, the difference of the two frequencies is cumulatively generated in an equally narrow beam. This nonlinear acoustic phenomenon was named as the parametric acoustic array and first discovered by Westervelt in the 1960s [1]. A widely studied application of the parametric acoustic array in air is the parametric array loudspeaker (PAL), which provides an efficient directional sound device that can be readily utilized in various applications of sound field control [2]–[4].

The sound process of the PAL is accurately described by the second order nonlinear acoustic equation. However, the second order nonlinear acoustic equation has only numerical solutions at high computational costs. A simplification to the second order nonlinear acoustic equation is provided by the Berktay's far-field solution [5]. It has been accepted as the fundamental acoustic model of most preprocessing methods of the PAL. Based on the Berktay's far-field solution, the sound pressure level of the PAL is proportional to the second derivative of the squared envelope function.

Conversely, the sound process of the PAL can be considered as the nonlinear distortion of the modulated ultrasonic carrier. Thus, harmonics of the desired audible wave are by-produced unavoidably. They are contributed to the unsatisfactory sound quality of the PAL. There have been a variety of preprocessing methods developed to reduce the harmonic distortion [6]–[9].

This work is supported by MEXT-Supported Program for the Strategic Research Foundation at Private University, 2013-2017.

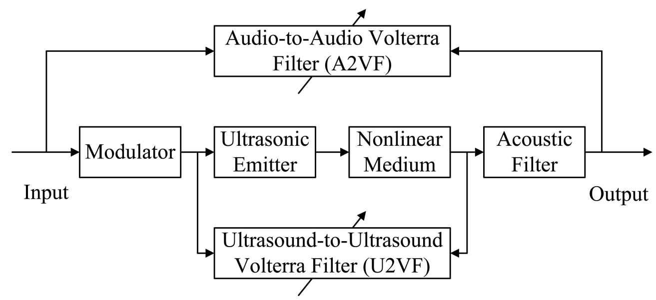


Fig. 1. Block diagrams of the parametric array loudspeaker and two types of Volterra filter identifications [16].

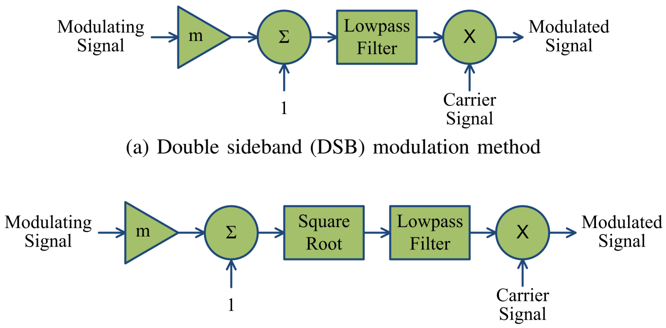
However, these preprocessing methods encounter difficulties when the applied environment does not fulfill the assumptions of the Berktay's far-field solution [10]. Hence, a preprocessing method able to adapt to the applied environment is desired.

To develop such an adaptive preprocessing method, the use of the Volterra filter to represent the sound process of the PAL can be traced back to 2002 [11]. It is noted in the past studies that the Volterra filter is always identified between the audio input and the audio output of the PAL [11]–[15]. This is thus referred to as the audio-to-audio Volterra filter (A2VF). It is similar to the linearization of the conventional loudspeaker. However, the A2VF does not take into account of the structure of the PAL. Therefore, the effectiveness of the A2VF is likely to be limited by the modulation method and modulation index.

Using the modulated signal of the PAL as the identification input, the ultrasound-to-ultrasound (U2VF) Volterra filter has recently been proposed [16]. Block diagrams of the A2VF and U2VF are shown in Fig. 1, together with the block diagram of the PAL. The U2VF includes no nonlinearity resultant from the modulator, where the preprocessing method is implemented. Therefore, an inverse system of the U2VF is expected to lead to the true linearization of the PAL, although such an inverse system is not readily available at this stage.

II. Parametric array loudspeaker

In the PAL, the audio input is modulated on an ultrasonic carrier. The modulated ultrasonic carrier is transmitted from the ultrasonic emitter at a sufficient pressure level to create the parametric acoustic array in air. The difference of the sideband and carrier frequencies in the modulated ultrasonic carrier wave provides a moderately distorted waveform of the



(a) Double sideband (DSB) modulation method

(b) Square root (SRT) modulation method

Fig. 2. Block diagrams of the (a) DSB and (b) SRT modulation methods [9].

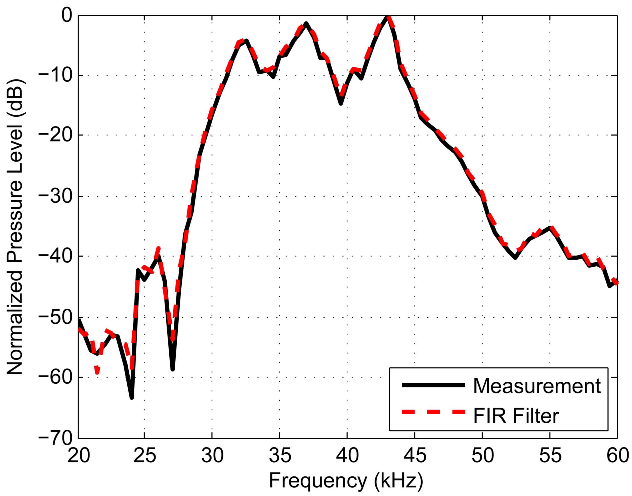


Fig. 3. Frequency response of the ultrasonic emitter.

audio input. This nonlinear acoustic process is also known as the self-demodulation effect [5]. Because the self-demodulated beam is as narrow as the beam of the ultrasonic carrier, the PAL is advantageous in creating a narrower sound beam when compared to other sound devices of the equivalent size.

A. Berktay's far-field solution

Although the Berktays far-field solution is a simple nonlinear acoustic equation, it is still the most widely applied model for designing the preprocessing method of the PAL. Based on the Berktays far-field solution, the self-demodulated pressure level of the PAL is given as

$$p_d\left(z\right) = \frac{\beta_0 {P_0}^2 S}{16\pi\rho_0 c_0^4 z \alpha_0} \frac{\partial^2}{\partial t^2} E^2\left(t - \frac{z}{c_0}\right), \tag{1}$$

where z is the distance between the ultrasonic emitter and the observation point; β_0 is the nonlinear coefficient of air; P_0 is the pressure level of the ultrasonic carrier wave; S is the size of the ultrasonic emitter; ρ_0 is the density of air; c_0 is the speed of sound in air; α_0 is the absorption coefficient at the carrier frequency; and $E\left(t\right)$ is the envelope function, which is provided by the preprocessed audio input. It is important to note that in Berktay's original derivation, the envelope function is assumed to have an unit amplitude [5].

B. Preprocessing methods

The disadvantage of the double sideband (DSB) modulation method has been elaborated in [6]. The envelope function of the DSB modulation method is given by

$$E_{DSB}\left(t\right) = \frac{1}{1+m} + \frac{m}{1+m} x\left(t\right), \tag{2}$$

where m is the modulation index; and $x\left(t\right)$ is the normalized audio input. Substituting (2) into (1) yields the conclusion that the self-demodulated pressure level and harmonic distortion of the DSB modulation method are both proportional to m.

To address the disadvantage of the DSB modulation method, the square root (SRT) modulation method has been proposed by Kamakura *et al.* to equalize the squared envelope function in the Berktay's far-field solution [7]. The envelope function of the SRT modulation method is given by

$$E_{SRT}\left(t\right) = \sqrt{\frac{1}{1+m} + \frac{m}{1+m} x\left(t\right)}. \tag{3}$$

Due to the square root operation, the SRT modulation method necessitate the ultrasonic emitter to possess an infinite bandwidth, which is not available off-the-shelf.

C. Ultrasonic emitter

It has been long recognized however often neglected that the frequency response of the ultrasonic emitter influences the actual performance of the preprocessing method. This is partially because that there is no particular parameter in the Berktay's far-field solution accounting for the frequency response of the ultrasonic emitter. One of the feasible simulation approach is to adopt a digital filter to approximate the frequency response of the ultrasonic emitter. The measured frequency response of the ultrasonic emitter and the frequency response of the finite impulse response (FIR) filter used in the simulation are plotted in Fig. 3 as an example.

The total harmonic distortion (THD) is defined as

$$\mathrm{THD}\left(f\right) = \sqrt{\frac{A^2\left(2f\right) + A^2\left(3f\right) + \ldots}{A^2\left(f\right)}}, \tag{4}$$

where f denotes the fundamental frequency and $A\left(nf\right)$ gives the amplitude of the frequency nf when $n \in \mathbb{N}$.

The total harmonic distortion (THD) curves using the DSB and SRT modulation methods are plotted in Figs. 4 and 5, respectively. The modulation index is set to 1.0. The simulation results with and without the FIR filter are apparently different. The simulation results without the FIR filter demonstrate the THD performance in theory, but the simulation results with the FIR filter agree with the measurement results. The importance of considering the frequency response of the ultrasonic emitter in the PAL is therefore confirmed.

III. Volterra Filter Identification

The Volterra filter is a popular signal processing approach to model a nonlinear system. In order to reduce the computational complexity, the Volterra filter is usually truncated at the second

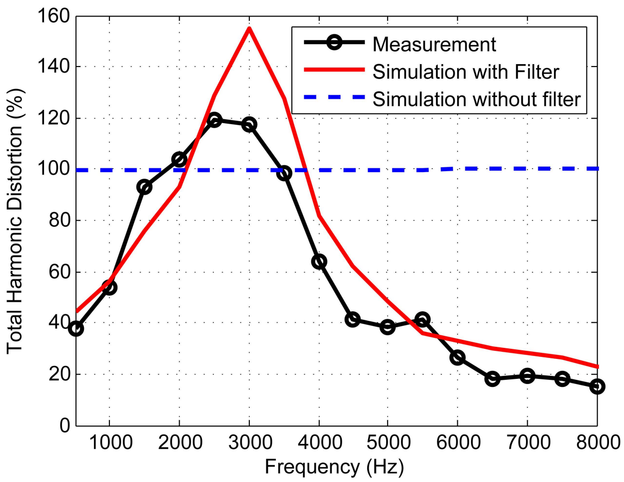


Fig. 4. THD performance of the DSB modulation method.

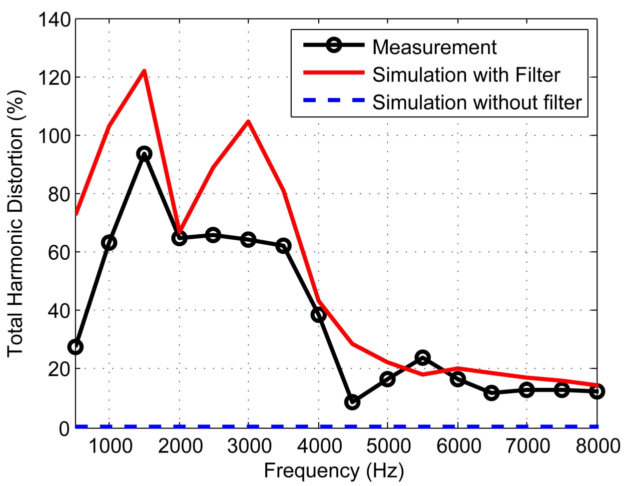


Fig. 5. THD performance of the SRT modulation method.

order when it is used to model the sound process of the PAL. The Volterra filter truncated at the second order is written as

$$y(n) = H_1[x(n)] + H_2[x(n)], \tag{5}$$

where $x(n)$ and $y(n)$ are the input and output of the nonlinear system; H_1 and H_2 are the first and second order Volterra operators, *i.e.*

$$H_1[x(n)] = \sum_{i=0}^{N-1} h_1(i)\, x(n-i) \tag{6}$$

and

$$H_2[x(n)] = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} h_2(i,j)\, x(n-i)\, x(n-j), \tag{7}$$

respectively. In (6) and (7), N is the memory length; h_1 and h_2 are the first and second order Volterra kernel coefficients.

In the adaptive Volterra filter identification, both the input and output of the nonlinear system are acquired continuously and stored in the vector form. Hence, (5) can be rewritten in a linear model as

$$y(n) = X^T H, \tag{8}$$

where

$$X = [x, x(n)\cdot x, x(n-1)\cdot x, \ldots, x(n-N+1)\cdot x]^T \tag{9}$$

and

$$\begin{aligned} H =& [h_1, h_2(0,0), h_2(0,1), \ldots, h_2(0,N-1), h_2(1,0), \\ & h_2(1,0), \ldots, h_2(1,N-1), \ldots, h_2(N-1,0), \\ & h_2(N-1,1), \ldots, h_2(N-1,N-1)]^T. \end{aligned} \tag{10}$$

Based on (8), the normalized least mean squares (NLMS) and sparse NLMS algorithms can be carried out to iteratively find the optimized solution of H. The update equations of the NLMS and sparse NLMS algorithms are given by

$$H_{k+1} = H_k + \alpha \frac{X_k e(k)}{X_k^T X_k} \tag{11}$$

and

$$H_{k+1} = H_k + \frac{1}{X_k^T X_k}\left[\alpha X_k e(k) - \beta \frac{\operatorname{sgn}(H_k)}{1 + |H_k|/\beta}\right], \tag{12}$$

respectively. In (11) and (12), α denotes the step size and β is the shrinking threshold. The coefficients below the shrinking threshold will shrink to 0 to create the sparsity [16].

IV. Experiment Results

The experiment has taken place in a sound proof room $(2.9 \times 3.1 \times 2.1\ \text{m}^3)$, where the microphone (B&K 4191L) is placed 3.0 m away from the ultrasonic emitter (Mitsubishi Electronic Engineering Company). The sampling frequency is set to 192 kHz. The carrier frequency is set to 40 kHz. The DSB modulation method is implemented with two modulation indexes of 1.0 and 0.5. A low-passed white noise and a sine sweep are generated for the A2VF identification and THD measurement of the PAL, respectively. The cut-off frequency of the white noise is 16 kHz. The sine sweep ranges from 1 kHz to 8 kHz. A band-passed white noise is generated for the U2VF identification, of which the cut-off frequencies are 24 kHz and 56 kHz. A postprocessing low-pass filter is applied to all the recorded signals except for the U2VF identification. The cut-off frequency of this low-pass filter is 18 kHz, in order to suppress high order harmonics that are inaudible and beyond our interest.

Firstly, the sparse NLMS algorithm identifies the first order Volterra kernel with a memory length of 3000 taps. This is to figure out the acoustic delay from the PAL to the microphone. Secondly, the recorded signal is shifted by this acoustic delay in order to shorten the memory length of the Volterra filter to 500 taps. Lastly, the identified A2VF and U2VF are adopted to evaluate the THD performance of the PAL. Since the white noises used for the A2VF and U2VF identifications have unit amplitudes, the sine sweep used for evaluating the THD performance has an unit amplitude too.

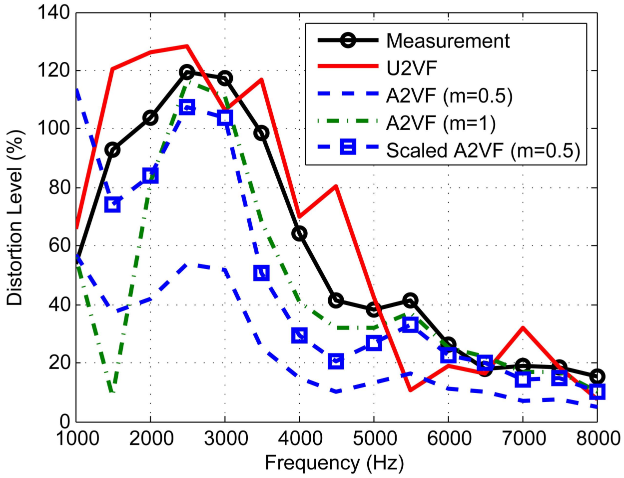


Fig. 6. THD performance of the DSB modulation method using the modulation index of $m = 1.0$.

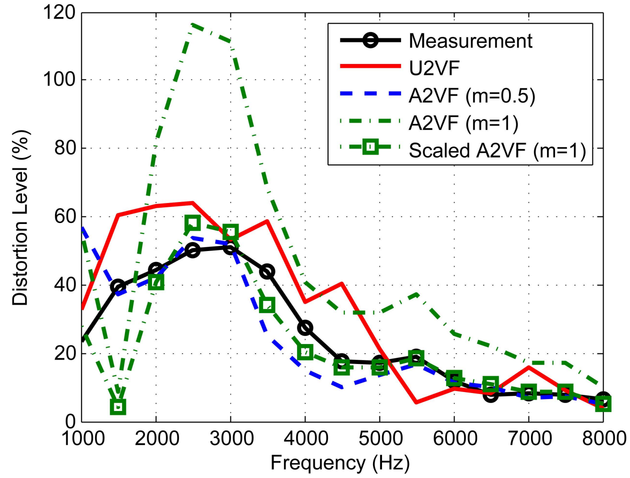


Fig. 7. THD performance of the DSB modulation method using the modulation index of $m = 0.5$.

The evaluation results of the A2VF and U2VF are plotted in Figs. 6 and 7, when the modulation index is chosen as 1.0 and 0.5, respectively. The measured THD curves are also provided for comparison. Figures 6 and 7 demonstrate that the U2VF is more robust than the A2VF, as the A2VF is unable to reflect the change of the modulation index. Nevertheless, the A2VF exhibits good accuracies when the modulation index is unchanged from the A2VF identification.

It is further attempted to scale the sine sweep according to the modulation index. For example, when the A2VF identified for the modulation index of 0.5 is used to evaluate the PAL using the modulation index of 1.0, the audio input is doubled. Similarly, when the A2VF identified for the modulation index of 1.0 is used to evaluate the PAL using the modulation index of 0.5, the audio input is scaled down by half. By scaling the sine sweep, the evaluation results of the A2VF can match the measurement results when the modulation index is different from the A2VF identification. This is mostly because that the DSB modulation method is linear. In the end, scaling the audio input is not recommended, as it leads to the wrong prediction of the amplitude of the audio output meanwhile.

V. Conclusions

The robustness of the U2VF to the change of the modulation method has been validated previously [16]. In this paper, the U2VF is examined under the change of the modulation index and compared with the A2VF for evaluating the THD performance of the PAL adopting the DSB modulation method. It is found that since the DSB modulation method is linear, scaling the audio input improves the accuracy of using the A2VF when the modulation index is changed. Still, the U2VF demonstrates the robustness to the modulation index, owing to the fact that it is independent from the modulator and preprocessing method of the PAL.

References

[1] P. J. Westervelt, "Parametric acoustic array," *J. Acoust. Soc. Amer.*, vol. 35, no. 4, pp. 535-537, 1963.

[2] C. Shi, Y. Kajiakawa, and W. S. Gan, "An overview of directivity control methods of the parametric array loudspeaker," *APSIPA Trans. Sig. Inform. Process.*, vol. 3, no. E20, pp. 1-12, 2015.

[3] C. Shi, H. Nomura, T. Kamakura, and W. S. Gan, "Spatial aliasing effects in a steerable parametric loudspeaker for stereophonic sound reproduction," *IEICE Trans. Fund. Electron. Commun. Computer Sci.*, vol. E97A, no.9, pp.1859-1866, 2014.

[4] K. Tanaka, C. Shi, and Y. Kajikawa, "Study on active noise control system using parametric array loudspeakers," in *Proc. 7th Forum Acusticum*, Krakow, Poland, 2014.

[5] H. O. Berktay, "Possible exploitation of nonlinear acoustics in underwater transmitting applications," *J. Sound Vib.*, vol. 2, no. 4, pp. 435-461, 1965.

[6] M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe, "The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design," *J. Acoust. Soc. Amer.*, vol. 73, no. 5, pp. 1013-1020, 1983.

[7] T. Kamakura, M. Yoneyama, and K. Ikegaya, "Developments of parametric loudspeaker for practical use," in *Proc. 10th Int. Symp. Nonlinear Acoust.*, Kobe, Japan, 1984.

[8] C. Shi, H. Mu, and W. S. Gan, "A psychoacoustical preprocessing technique for virtual bass enhancement of the parametric loudspeaker," in *Proc. 38th Int. Conf. Acoust. Speech Sig. Process.*, Vancouver, Canada, 2013.

[9] C. Shi and W. S. Gan, "A preprocessing method to increase high frequency response of a parametric loudspeaker," in *Proc. 2013 APSIPA Annu. Summit Conf.*, Kaohsiung, Taiwan, 2013.

[10] C. Shi and Y. Kajikawa, "A comparative study of preprocessing methods in the parametric loudspeaker," in *Proc. 2014 APSIPA Annu. Summit Conf.*, Siem Reap, Cambodia, 2014.

[11] C. M. Lee, J. Yang, W. S. Gan, and M. H. Er, "Modeling nonlinearity of air with Volterra kernels for use in a parametric array loudspeaker," in *Proc. 112th AES Conv.*, Munich, Germany, 2002.

[12] W. Ji and W. S. Gan, "Identification of a parametric loudspeaker system using an adaptive Volterra filter," *Appl. Acoust.*, vol. 73, no. 12, pp. 1251-1262, 2012.

[13] O. Hoshuyama and T. Okayasu, "Decimation in Volterra filter for distortion reduction of ultrasonic parametric loudspeaker," in *Proc. 2014 IEICE Soc. Conf.*, Tokushima, Japan, 2014.

[14] Y. Hatano, C. Shi, and Y. Kajikawa, "A study on linearization of nonlinear distortions in parametric array loudspeakers," in *Proc. 2014 Int. Workshop Smart Inf. Media Syst.*, Ho Chi Minh City, Vietnam, 2014.

[15] Y. Mu, P. Ji, W. Ji, M. Wu, and J. Yang, "Modeling and compensation for the distortion of parametric loudspeakers using a one-dimension Volterra filter," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 12, pp. 2169-2181, 2014.

[16] C. Shi and Y. Kajikawa, "Identification of the parametric array loudspeaker with a Volterra filter using the sparse nlms algorithm," in *Proc. 40th Int. Conf. Acoust. Speech Signal Process.*, Brisbane, Australia, 2015.